

Algebraic analysis of two-grid methods: The nonsymmetric case

Yvan Notay^{*,†}

*Service de Métrologie Nucléaire, Université Libre de Bruxelles, C.P. 165/84,
50, Av. F.D. Roosevelt, B-1050 Brussels, Belgium*

SUMMARY

Two-grid methods constitute the building blocks of multigrid methods, which are among the most efficient solution techniques for solving large sparse systems of linear equations. In this paper, an analysis is developed that does not require any symmetry property. Several equivalent expressions are provided that characterize *all* eigenvalues of the iteration matrix. In the symmetric positive-definite (SPD) case, these expressions reproduce the sharp two-grid convergence estimate obtained by Falgout, Vassilevski and Zikatanov (*Numer. Linear Algebra Appl.* 2005; **12**:471–494), and also previous algebraic bounds, which can be seen as corollaries of this estimate. These results allow to measure the convergence by checking ‘approximation properties’. In this work, proper extensions of the latter to the nonsymmetric case are presented. Sometimes approximation properties for the SPD case are summarized in loose terms; e.g.: *Interpolation must be able to approximate an eigenvector with error bound proportional to the size of the eigenvalue* (*SIAM J. Sci. Comp.* 2000; **22**:1570–1592). It is shown that this can be applied to nonsymmetric problems too, understanding ‘size’ as ‘modulus’. Eventually, an analysis is developed, for the nonsymmetric case, of the theoretical foundations of ‘compatible relaxation’, according to which a Fine/Coarse partitioning may be checked and possibly improved. Copyright © 2009 John Wiley & Sons, Ltd.

Received 6 June 2008; Revised 3 March 2009; Accepted 19 March 2009

KEY WORDS: multigrid; convergence analysis; linear systems; approximation property; convergence measure; compatible relaxation

1. INTRODUCTION

We consider two-grid methods for solving $n \times n$ systems of linear equations

$$\mathbf{A}\mathbf{u} = \mathbf{b} \tag{1}$$

*Correspondence to: Yvan Notay, Service de Métrologie Nucléaire, Université Libre de Bruxelles, C.P. 165/84, 50, Av. F.D. Roosevelt, B-1050 Brussels, Belgium.

†E-mail: ynotay@ulb.ac.be

Contract/grant sponsor: Belgian FNRS (‘Directeur de recherches’)

A two-grid method (e.g. [1]) combines the action of a *smoother*, often a simple iterative method such as Gauss–Seidel, and a *coarse-grid correction*, which involves solving a smaller problem on a coarser grid. More precisely, letting M be the smoother, the iteration matrix associated to one smoothing step is

$$T_s = I - M^{-1}A \quad (2)$$

On the other hand, letting n_c be the number of coarse variables, the coarse-grid correction involves a *restriction* matrix R of size $n_c \times n$, a *prolongation matrix* P of size $n \times n_c$, and a coarse-grid matrix A_c of size $n_c \times n_c$, which we assume in this work of Galerkin type; that is

$$A_c = RAP \quad (3)$$

The iteration matrix associated to one coarse-grid correction step is then

$$T_c = I - PA_c^{-1}RA \quad (4)$$

and the global iteration matrix for the two-grid method with v_1 pre- and v_2 post-smoothing steps is

$$T = T_s^{v_2} T_c T_s^{v_1} \quad (5)$$

As is well known, the asymptotic convergence factor is equal to $\rho(T)$ (the spectral radius of T); i.e. fast convergence is obtained if and only if $\rho(T)$ is significantly below one. Note that such a result has to be obtained despite $\rho(T_c) = 1$ (since T_c is a projector) whereas, in common applications, $\rho(T_s)$ is below 1 but close to 1 (otherwise the smoother alone would converge fast enough).

Several analyses have been developed for such methods. Abstract theories (e.g. [2, 3]) are restricted to discretized partial differential equations (PDEs) on a regularly refined grid, and allow to obtain only qualitative results. Fourier analysis (e.g. [1]) may be used when A , R and P correspond to a constant stencil applied on a regular grid. Then, all eigenvalues of T are easily computed, and sometimes analytical bounds may be derived; see, e.g. [4]. Only a few such results (e.g. [5, 6]) have been obtained for nonsymmetric problems, probably because the eigenvalues are complex and therefore difficult to bound, even when they are known through some analytical expression.

Besides, ‘algebraic’ analyses have been developed [7–12]. These may be applied to constant stencil problems, and, although they are not as accurate as Fourier analysis, they may offer shorter paths to analytical bounds. The intended application is, however, algebraic multigrid (AMG) methods. In these methods (e.g. [7, 11, 12]), the prolongation P is computed from the system matrix A by means of a so-called ‘coarsening’ algorithm. Then, all other analyses fail because P does not have, in general, a completely regular pattern even when A corresponds, say, to a constant stencil. Note that these algebraic bounds may be used to analyze or validate a given scheme, but also as guidelines in the development of new methods; see, e.g. [13–15] for examples.

These algebraic analyses share some common features. They apply only to the symmetric positive-definite (SPD) case; that is, they require A SPD, $R = P^T$ and a symmetric smoothing scheme (see Section 2 for a precise definition). In this framework, they lead to a bound of the form $\rho(T) \leq 1 - 1/K$, where K is a constant, which mainly depends on an *approximation property*

$$\mathbf{v}^*(I - Q)^T Y (I - Q) \mathbf{v} \leq c \mathbf{v}^* A \mathbf{v} \quad \text{for all } \mathbf{v} \in \mathcal{C}^n \quad (6)$$

for some nontrivial (not too large) constant c ; in this expression, and depending on which of the results from [7–12] is more precisely concerned, Y is either the diagonal of A or an SPD matrix related to the smoother, and Q a projector onto the range of P , either prescribed or with some

freedom left in. The analysis also yields a sharp estimate for some specific choice of Y and Q , see [10] or Section 2 for details.

In this paper, we extend these analyses to the nonsymmetric case. This extension also concerns Hackbush analysis for the SPD case [2, p. 151], although it does not fit exactly in the above framework. One outcome is that approximation properties as stated above have also some meaning for general matrices, provided that one takes the modulus of $\mathbf{v}^* A \mathbf{v}$ on the right-hand side. Therefore, in the nonsymmetric case too, the convergence is strongly related to the quality of the interpolation for ‘low-energy modes’, to be understood as ‘modes for which $|\mathbf{v}^* A \mathbf{v}|$ is small’.

Note, however, that the latter statement is an over-simplification of the truth; it relies on the further assumption that the low-energy modes (or the modes for which $|\mathbf{v}^* A \mathbf{v}|$ is small) are poorly treated by relaxation. This is the case for many of the simple smoothers commonly used, but is not true for certain cases, such as the distributive or coupled smoothers commonly used for discretized systems of PDEs. The disadvantage of considering classical smoothing and approximation properties is that this distinction cannot be made in that theory. In fact, our main result (Theorem 2.1) does not suffer from such a drawback, but, so far, it is unclear whether it can be used in a practical setting that goes beyond the classical approximation properties.

When A is SPD and Y is the diagonal of A , (6) is the approximation property considered in the ‘standard’ analysis of AMG methods, as developed in [7, 11, 12]. Here, we also show that the constant in this property is directly related to the convergence rate of the two-grid method with only one damped Jacobi pre- or post-smoothing iteration. That is, requiring this constant to be moderate implies some minimal requirement on the convergence of this (over) simplified two-grid method.

Another outcome of our study is related to so-called ‘compatible relaxation’ [9, 16, 17]. Often, AMG methods are based on a two-step procedure: one first defines a Fine/Coarse partitioning by inspecting the matrix graph; next, P is built using interpolation rules for fine-grid variables. In this context, compatible relaxation allows to check the quality of the obtained Fine/Coarse partitioning. It is also possible to improve it, or even to construct a partitioning from scratch, see [17] for a thorough study, and [18] for a related strategy. In this work, we consider the extension to nonsymmetric problems of the theoretical foundations of this approach [9].

Eventually, it is worth mentioning here the works by Mense and Nabben, who recently developed an analysis of multilevel methods for nonsymmetric matrices [19, 20]. These works are complementary to ours because they address methods of so-called AMLI type, that are based on a block incomplete factorization of the matrix partitioned in the hierarchical form (see [21, 22] for a comparative discussion of different multilevel schemes). So far, the results obtained with this approach, which is based on the classical convergent splitting theory, are also essentially qualitative and restricted to M -matrices.

The remainder of this paper is organized as follows. Section 2 contains our main theoretical results and a related discussion of the approximation properties. Compatible relaxation is considered in Section 3.

1.1. General terminology and notation

We note I as the identity matrix; when the dimensions are not obvious from the context, I_m denotes more specifically the $m \times m$ identity matrix. When matrices are written in block form, the symbol ‘*’ stands for blocks whose expressions are unimportant for the related discussion or the result being proved.

For any matrix C , C^* denotes its conjugate transpose. If C is square, $\sigma(C)$ is its spectrum, $\rho(C)$ is its spectral radius (that is, its largest eigenvalue in modulus) and $\|C\| = \sqrt{\rho(C^*C)}$ is the usual 2–norm; if the eigenvalues of C are real, $\lambda_{\min}(C)$ and $\lambda_{\max}(C)$ denote, respectively, the smallest and the largest eigenvalue of C . For any Hermitian positive-definite matrix B , $\|C\|_B = \|B^{1/2}CB^{-1/2}\|$ is the B –norm (often referred to as energy norm).

When we write that an $m \times m$ matrix C has eigenvalues $\lambda_1, \dots, \lambda_m$, it is always counting multiplicities; that is, its characteristic polynomial is $(\lambda_1 - \lambda) \cdots (\lambda_m - \lambda)$. When we write that two matrices have the same eigenvalues, it is also always counting multiplicities; that is, they have the same characteristic polynomial.

2. MAIN RESULTS AND APPROXIMATION PROPERTIES

Our analysis is slightly more general than stated in the introduction where, for the sake of simplicity, we consider a single nonsingular smoother. In fact, we allow different smoothers in pre- and post-smoothing steps (for instance, forward and backward Gauss–Seidel). We even allow singular smoothers; that is, we write

$$T_{s_1} = I - M_1A \quad \text{and} \quad T_{s_2} = I - M_2A$$

as the iteration matrices for pre- and post-smoothing, respectively, where M_1, M_2 may be singular. We, however, assume that there exists a nonsingular matrix X such that

$$I - X^{-1}A = (I - M_1A)^{v_1}(I - M_2A)^{v_2} \quad (7)$$

Hence, M_1A and M_2A may not have a common singular mode. The matrix X may be seen as the equivalent ‘global’ smoother, which brings in one step the same effect as post-smoothing followed by pre-smoothing. Note that if $M_1 = M_2 = M^{-1}$ are nonsingular, there holds

$$X^{-1}M = (I - T_s^v)(I - T_s)^{-1} = I + T_s + \cdots + T_s^{v-1}$$

where $v = v_1 + v_2$. It can be seen that the right-hand side matrix has no eigenvalue close to zero if T_s has no eigenvalue close to $e^{i2\pi k/v}$ for $k = 1, \dots, v-1$. In most cases, this is ensured by a proper scaling of M . Hence, the assumption that X exists and is nonsingular is not really restrictive.

Now, our main result is given in the following theorem. Note that we do not assume either explicitly or implicitly that the matrices are real; that is, all matrices referred in Theorem 2.1 may be complex valued. In this theorem, we assume that $A_c = RAP$ and $X_c = RXP$ are nonsingular. Because A and X are possibly indefinite, this is indeed not guaranteed by the assumption that A and X are nonsingular and that P and R have full rank. Such a guarantee however holds for certain class of matrices, for instance those for which the Hermitian parts $\frac{1}{2}(A + A^*)$ and $\frac{1}{2}(X + X^*)$ are positive definite.

Theorem 2.1

Let A be a nonsingular $n \times n$ matrix. Let P be an $n \times n_c$ matrix of rank n_c and let R an $n_c \times n$ matrix of rank n_c , such that $A_c = RAP$ is nonsingular. Let M_1, M_2 be $n \times n$ matrices and v_1, v_2 be nonnegative integers, such that $(I - M_1A)^{v_1}(I - M_2A)^{v_2} - I$ is nonsingular. Let X be the matrix such that

$$I - X^{-1}A = (I - M_1A)^{v_1}(I - M_2A)^{v_2}$$

and assume that

$$X_c = RXP$$

is nonsingular.

The matrix

$$A^{-1}X(I - \pi_X) \quad (8)$$

where

$$\pi_X = PX_c^{-1}RX$$

has eigenvalue 0 with multiplicity n_c and $n - n_c$ nonzero eigenvalues. Letting $\mu_1, \dots, \mu_{n-n_c}$ be these nonzero eigenvalues, the following propositions hold:

1. The iteration matrix

$$T = (I - M_2A)^{v_2}(I - PA_c^{-1}RA)(I - M_1A)^{v_1} \quad (9)$$

has eigenvalues $1 - \mu_1^{-1}, \dots, 1 - \mu_{n-n_c}^{-1}$, plus n_c times the eigenvalue 0.

2. For any $(n - n_c) \times n$ matrix Z and any $n \times (n - n_c)$ matrix S such that the matrices

$$\begin{pmatrix} R \\ Z \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} P & S \end{pmatrix}$$

are nonsingular, $(ZAS) - (ZAP)A_c^{-1}(RAS)$ is nonsingular and the matrix

$$((ZAS) - (ZAP)A_c^{-1}(RAS))^{-1}ZX(I - \pi_X)S \quad (10)$$

has eigenvalues $\mu_1, \dots, \mu_{n-n_c}$.

3. For any $n \times n_c$ matrix \tilde{P} and any $n_c \times n$ matrix \tilde{R} , the matrix

$$A^{-1}(I - \tilde{P}\tilde{R})X(I - \pi_X)(I - P\tilde{R}) \quad (11)$$

has the same eigenvalues as (8); that is, $\mu_1, \dots, \mu_{n-n_c}$, plus n_c times the eigenvalue 0.

4. The matrices

$$(I - \pi_A)X^{-1}A \quad (12)$$

$$X^{-1}A(I - \pi_A) \quad (13)$$

$$(I - \pi_A)X^{-1}A(I - \pi_A) \quad (14)$$

where

$$\pi_A = PA_c^{-1}RA$$

have eigenvalues $\mu_1^{-1}, \dots, \mu_{n-n_c}^{-1}$, plus n_c times the eigenvalue 0.

5. For all μ_i , $i = 1, \dots, n - n_c$, there exist some $\mathbf{z}_i \in \mathcal{C}^n$ such that $R\mathbf{z}_i = 0$ and

$$\mathbf{v}^* A^{-1} \mathbf{z}_i = \mu_i \mathbf{v}^* X^{-1} \mathbf{z}_i \quad \text{for all } \mathbf{v} \in \mathcal{C}^n : P^* \mathbf{v} = 0 \quad (15)$$

6. If, in addition, $R = P^*$ and there is no $\mathbf{v} \in \mathcal{C}^n$ such that $P^*\mathbf{v} = 0$ and $\mathbf{v}^*A^{-1}\mathbf{v} = \mathbf{v}^*X^{-1}\mathbf{v} = 0$, then, for $i = 1, \dots, n_{n-n_c}$

$$\mu_i \in \left\{ \frac{\mathbf{v}^*A^{-1}\mathbf{v}}{\mathbf{v}^*X^{-1}\mathbf{v}} \mid \mathbf{v} \in \mathcal{C}^n \setminus \{0\}, P^*\mathbf{v} = 0, \mathbf{v}^*A^{-1}\mathbf{v} \neq 0 \text{ and } \mathbf{v}^*X^{-1}\mathbf{v} \neq 0 \right\} \quad (16)$$

Proof

We shall conduct the proof as follows. Let η_1, \dots, η_n be the eigenvalues of

$$\pi_X + A^{-1}X(I - \pi_X)$$

We shall prove that: (a) all these eigenvalues are nonzero, and the eigenvalues of T are $1 - \eta_1^{-1}, \dots, 1 - \eta_n^{-1}$; (b) (at least) n_c of these η_i are equal to 1, and the other η_i are the $n - n_c$ eigenvalues of the matrix (10), where $(ZAS) - (ZAP)A_c^{-1}(RAS)$ is nonsingular; (c) the eigenvalues of the matrix (11) are all the eigenvalues of the matrix (10), plus n_c times the eigenvalue 0.

Since the matrix (8) is the matrix (11) when $\tilde{P} = 0$ and $\tilde{R} = 0$, (c) implies that it has effectively n_c times the eigenvalue 0, whereas the other eigenvalues (that is, the μ_i) are equal to the ‘other’ η_i mentioned above, which are shown in (a) to be nonzero. One then readily finds that altogether (a), (b), (c) imply our claims about the eigenvalues of the matrix (8) and statements 1–3 of the theorem. Statements 4–6 will be proved subsequently.

Thus we first prove (a). We use the fact that for any square matrices C_1, C_2 , the products C_1C_2 and C_2C_1 have the same eigenvalues. This fact is obvious when at least one of the matrices is nonsingular (there exists a similarity transformation), but is less well known in the general case; at any rate, we give a proof in Lemma A.1 of the appendix. It follows that T has the same eigenvalues as

$$\tilde{T} = (I - X^{-1}A)(I - PA_c^{-1}RA)$$

Consider then

$$B = \begin{pmatrix} I & \\ -R & I \end{pmatrix} \begin{pmatrix} X & \\ & A_c \end{pmatrix} \begin{pmatrix} I & -(I - X^{-1}A)P \\ & I \end{pmatrix} \quad (17)$$

$$\begin{aligned} &= \begin{pmatrix} X & -(X - A)P \\ -RX & X_c \end{pmatrix} \\ &= \begin{pmatrix} I & -(X - A)PX_c^{-1} \\ & I \end{pmatrix} \begin{pmatrix} X - (X - A)PX_c^{-1}RX & \\ & X_c \end{pmatrix} \begin{pmatrix} I & \\ -X_c^{-1}RX & I \end{pmatrix} \quad (18) \end{aligned}$$

Note that B is nonsingular and therefore so is $X - (X - A)PX_c^{-1}RX = X - (X - A)\pi_X$. Further, (18) gives

$$\begin{aligned} B^{-1} &= \begin{pmatrix} I & \\ X_c^{-1}RX & I \end{pmatrix} \begin{pmatrix} (X - (X - A)\pi_X)^{-1} & \\ & X_c^{-1} \end{pmatrix} \begin{pmatrix} I & (X - A)PX_c^{-1} \\ & I \end{pmatrix} \\ &= \begin{pmatrix} (X - (X - A)\pi_X)^{-1} & * \\ & * \end{pmatrix} \end{aligned}$$

whereas (17) yields

$$\begin{aligned} B^{-1} &= \begin{pmatrix} I & (I - X^{-1}A)P \\ & I \end{pmatrix} \begin{pmatrix} X^{-1} & \\ & A_c^{-1} \end{pmatrix} \begin{pmatrix} I & \\ R & I \end{pmatrix} \\ &= \begin{pmatrix} X^{-1} + (I - X^{-1}A)P A_c^{-1}R & * \\ & * \end{pmatrix} \\ &= \begin{pmatrix} (I - \tilde{T})A^{-1} & * \\ & * \end{pmatrix} \end{aligned}$$

Therefore, $I - \tilde{T}$ is nonsingular and

$$(I - \tilde{T})^{-1} = A^{-1}(X - (X - A)\pi_X) = \pi_X + A^{-1}X(I - \pi_X)$$

proving (a) (the η_i are nonzero because $I - \tilde{T}$ is nonsingular).

To pursue, consider the matrices

$$F = \begin{pmatrix} R \\ Z \end{pmatrix}, \quad G = (P \ S)$$

and let

$$F^{-1} = (\hat{P} \ W), \quad G^{-1} = \begin{pmatrix} \hat{R} \\ V \end{pmatrix}$$

where \hat{P} , W , \hat{R} , V are, respectively, $n \times n_c$, $n \times (n - n_c)$, $n_c \times n$, $(n - n_c) \times n$. From $FF^{-1} = I$ and $G^{-1}G = I$, one has

$$R\hat{P} = I_{n_c}, \quad \hat{R}P = I_{n_c} \quad (19)$$

$$ZW = I_{n-n_c}, \quad VS = I_{n-n_c} \quad (20)$$

$$RW = 0, \quad \hat{R}S = 0 \quad (21)$$

$$Z\hat{P} = 0, \quad VP = 0 \quad (22)$$

whereas $F^{-1}F = I$ and $GG^{-1} = I$ give

$$\hat{P}R + WZ = I_n, \quad P\hat{R} + SV = I_n \quad (23)$$

Now, $\pi_X + A^{-1}X(I - \pi_X)$ has the same eigenvalues as

$$\begin{aligned} G^{-1}(\pi_X + A^{-1}X(I - \pi_X))G &= \begin{pmatrix} \hat{R} \\ V \end{pmatrix} (\pi_X + A^{-1}X(I - \pi_X))(P \ S) \\ &= \begin{pmatrix} \hat{R} \\ V \end{pmatrix} (P \ (\pi_X + A^{-1}X(I - \pi_X))S) \\ &= \begin{pmatrix} I & * \\ 0 & VA^{-1}X(I - \pi_X)S \end{pmatrix} \end{aligned} \quad (24)$$

Further,

$$\begin{aligned} VA^{-1}X(I-\pi_X)S &= VA^{-1}(\widehat{P}R+WZ)X(I-\pi_X)S \\ &= (VA^{-1}W)(ZX(I-\pi_X)S) \end{aligned} \quad (25)$$

Consider then

$$G^{-1}A^{-1}F^{-1} = \begin{pmatrix} * & * \\ * & VA^{-1}W \end{pmatrix} \quad (26)$$

Letting $C = (ZAS) - (ZAP)A_c^{-1}(RAS)$, one finds that

$$\begin{aligned} G^{-1}A^{-1}F^{-1} &= (FAG)^{-1} \\ &= \left(\begin{pmatrix} I & \\ (ZAP)A_c^{-1} & I \end{pmatrix} \begin{pmatrix} A_c & \\ & C \end{pmatrix} \begin{pmatrix} I & A_c^{-1}(RAS) \\ & I \end{pmatrix} \right)^{-1} \\ &= \begin{pmatrix} I & -A_c^{-1}(RAS) \\ & I \end{pmatrix} \begin{pmatrix} A_c^{-1} & \\ & C^{-1} \end{pmatrix} \begin{pmatrix} I & \\ -(ZAP)A_c^{-1} & I \end{pmatrix} \\ &= \begin{pmatrix} * & * \\ * & C^{-1} \end{pmatrix} \end{aligned}$$

showing with (26) that C is nonsingular and such that $C^{-1} = VA^{-1}W$, which, together with (24), (25), proves (b).

Further, from (24)–(26), the eigenvalues of (10), since they are nonzero, are also the nonzero eigenvalues of

$$G^{-1}A^{-1}F^{-1} \begin{pmatrix} 0 & 0 \\ 0 & ZX(I-\pi_X)S \end{pmatrix}$$

Since $X(I-\pi_X)P=0$ and $RX(I-\pi_X)=0$, one finds

$$\begin{aligned} G^{-1}A^{-1}F^{-1} \begin{pmatrix} 0 & 0 \\ 0 & ZX(I-\pi_X)S \end{pmatrix} &= G^{-1}A^{-1}F^{-1} \begin{pmatrix} R \\ Z \end{pmatrix} X(I-\pi_X)(P \ S) \\ &= G^{-1}A^{-1}X(I-\pi_X)G \\ &= G^{-1}A^{-1}(I-\widetilde{P}R)X(I-\pi_X)(I-P\widetilde{R})G \end{aligned}$$

hence (c).

We are therefore left with the proof of statements 4–6. Let us choose S such that its $n-n_c$ columns are linearly independent and orthogonal to the n_c columns of $(RA)^*$ (i.e. $RAS=0$). Then, $(P \ S)$ is necessarily nonsingular since

$$\begin{aligned} (P \ S) \begin{pmatrix} \mathbf{v}_p \\ \mathbf{v}_s \end{pmatrix} = 0 &\Rightarrow (RAP \ RAS) \begin{pmatrix} \mathbf{v}_p \\ \mathbf{v}_s \end{pmatrix} = 0 \\ &\Rightarrow A_c \mathbf{v}_p = 0 \end{aligned}$$

$$\Rightarrow \mathbf{v}_p = 0$$

$$\Rightarrow S\mathbf{v}_s = 0$$

$$\Rightarrow \mathbf{v}_s = 0$$

Further, the matrix \widehat{R} satisfying (19), (21) is then $\widehat{R} = A_c^{-1}RA$, implying with (23) that $SV = I - P\widehat{R} = I - \pi_A$. Hence

$$\begin{aligned} (VX^{-1}AS)(VA^{-1}X(I - \pi_X)S) &= VX^{-1}A(I - \pi_A)A^{-1}X(I - \pi_X)S \\ &= V(I - \pi_X)S \\ &= VS \\ &= I_{n-n_c} \end{aligned}$$

showing that $(VX^{-1}AS)$ is the inverse of the matrix appearing at the bottom right block of (24). We have seen above that the latter matrix has μ_i as eigenvalues, hence the eigenvalues of $(VX^{-1}AS)$ are $\mu_1^{-1}, \dots, \mu_{n-n_c}^{-1}$. The proof of statement 4 is then concluded observing that

$$\begin{aligned} G^{-1}(I - \pi_A)X^{-1}A(I - \pi_A)G &= \begin{pmatrix} \widehat{R} \\ V \end{pmatrix} (I - P\widehat{R})X^{-1}A(I - P\widehat{R})(P \ S) \\ &= \begin{pmatrix} 0 & 0 \\ 0 & VX^{-1}AS \end{pmatrix} \end{aligned}$$

whereas Lemma A.1 of the Appendix shows that the matrices (12), (13) have the same eigenvalues as (14) (using $(I - \pi_A)^2 = I - \pi_A$).

To prove statement 5, let \mathbf{w}_i be an eigenvector of matrix (12) corresponding to a nonzero eigenvalue μ_i^{-1} , and set $\mathbf{z}_i = A\mathbf{w}_i$. Since $RA(I - \pi_A) = 0$, from

$$\mu_i(I - \pi_A)X^{-1}\mathbf{z}_i = A^{-1}\mathbf{z}_i \quad (27)$$

one sees, multiplying both sides on the left by RA , that $R\mathbf{z}_i = 0$. Statement 5 then straightforwardly follows since $P^*\mathbf{v} = 0$ entails $\mathbf{v}^*\pi_A = 0$.

Eventually, statement 6 is a straightforward corollary of (27), noting that, since the μ_i are nonzero (and finite), for any \mathbf{z}_i satisfying (27), $\mathbf{z}_i^*A^{-1}\mathbf{z}_i$ and $\mathbf{z}_i^*X^{-1}\mathbf{z}_i$ are either both nonzero or both equal to zero. \square

2.1. General discussion

Consider the matrix (8)

$$A^{-1}X(I - \pi_X) = A^{-1}(X - XP(RXP)^{-1}RX) \quad (28)$$

It has the same eigenvalues as $X(I - \pi_X)A^{-1}$ whose eigenvalues are, in turn, the complex conjugate of the eigenvalues of its conjugate transpose

$$(A^*)^{-1}(X^* - X^*R^*(P^*X^*R^*)^{-1}P^*X^*) \quad (29)$$

Comparing with (28), one sees that R, P form a good pair of restriction and prolongation for A with given global smoother X if and only if P^*, R^* form a good pair of restriction and prolongation for A^* with global smoother X^* . This observation is in line with the analysis in [23], where it is suggested to choose the restriction as the adjoint of the prolongation one would construct for A^* . This suggestion traces back to [24] (see, also [25]) where considering convection–diffusion problems, it is however combined with a prolongation computed from $\frac{1}{2}(A + A^*)$. At first sight, our results seem to indicate that if such a prolongation is a good one, then the restriction should also be computed from $\frac{1}{2}(A + A^*)$, leading to select $R = P^*$. This is partly confirmed by the analysis below, which shows that when $R = P^*$, it is sensible to construct both of them from $\frac{1}{2}(A + A^*)$. The latter choice is however certainly not *always* the best one, as discussed in [23], based on the example of the discrete problem arising from $i\Delta u = f$: in this case, the system matrix has a zero Hermitian part although the problem is easy to solve with multigrid methods. In fact, additional developments are needed to see whether an analysis based on Theorem 2.1 allows or not to improve the strategies proposed so far.

It is also worth mentioning here the recent results in [26]. There, an analysis based on an equivalent symmetric problem suggests to construct the prolongation (resp. the restriction) such that its range (resp. the range of its transpose) contains good approximations on the right (resp. left) near kernel modes. Here again, this is in line with the conclusions that can be drawn from Theorem 2.1. In [26], near kernel modes are defined from the singular value decomposition of A ; i.e. they correspond to the right or left singular vectors. Considering (28) or (29), one may think that the key role is played by the right or left eigenvectors, instead. However, as discussed below, it is above all important to bound $\max_i |\mu_i|$. This can be done by bounding the norm of these matrices, and, developing the analysis in this way, it seems indeed that the singular vectors of A are important too. Here also additional developments would be welcome, but we are raising questions that lie outside the scope of the present paper.

2.2. Symmetric restriction: $R = P^*$

Despite the above remarks, it is important to analyze the case where $R = P^*$. This is indeed the standard setting for geometric multigrid methods and, so far, also the choice most frequently made in practice for AMG methods (e.g. [27, 28]).

Let us first comment on the ‘SPD case’; that is, the case for which the analyses in [7–12] apply. These analyses require not only A real and SPD with $R = P^T$, but also a symmetric and convergent smoothing scheme: $M_2 = M_1^T = M^{-1}$, $v_1 = v_2$ and $\rho(I - M^{-1}A) < 1$ (sometimes, but not always, M requires SPD too).

With these assumptions, simple algebraic arguments show that the eigenvalues of the iteration matrix are real and in interval $[0, 1)$. Hence the μ_i are real positive and the analyses in [7–12] prove bounds of the form[‡] $\rho(T) \leq 1 - 1/K$, where K is in fact an upper bound on $\max_i \mu_i$.

Now, what can be said, more generally, about the eigenvalues of T knowing only A and X^{-1} ? The following corollary answers this question.

[‡]Sometimes the results are stated in the form $\|T_{s_2} T_c\|_A \leq \sqrt{1 - 1/K}$, but this is equivalent to $\rho(T) \leq 1 - 1/K$ because, under the assumptions of the SPD case, $\rho(T) = \|T_{s_2} T_c\|_A^2$, see, e.g. [21, p. 442] for a proof.

Corollary 2.1

If, in addition to the assumptions of Theorem 2.1, $R = P^*$ and there is no $\mathbf{v} \in \mathcal{C}^n$ such that $\mathbf{v}^* A^{-1} \mathbf{v} = \mathbf{v}^* X^{-1} \mathbf{v} = 0$, then,

$$\sigma(T) \subset \left\{ 1 - \frac{\mathbf{v}^* X^{-1} \mathbf{v}}{\mathbf{v}^* A^{-1} \mathbf{v}} \mid \mathbf{v} \in \mathcal{C}^n \setminus \{0\}, \mathbf{v}^* A^{-1} \mathbf{v} \neq 0 \text{ and } \mathbf{v}^* X^{-1} \mathbf{v} \neq 0 \right\} \cup \{0\} \quad (30)$$

In particular, if A, X are Hermitian positive definite, then the μ_i are real positive

$$\sigma(T) \subset [1 - \lambda_{\max}(X^{-1}A), 1 - \lambda_{\min}(X^{-1}A)] \cup \{0\} \quad (31)$$

and

$$\rho(T) \leq \max \left(\lambda_{\max}(X^{-1}A) - 1, 1 - \frac{1}{\max_i \mu_i} \right) \quad (32)$$

On the other hand, if X is Hermitian positive definite and if

$$\|\alpha I - X^{-1}A\|_X \leq \alpha \quad (33)$$

for some positive α , then

$$\lambda \in \sigma(T) \Rightarrow |1 - \alpha - \lambda| \leq \alpha \quad (34)$$

Proof

Relation (30) straightforwardly follows from statement 6 of Theorem 2.1. When A, X are Hermitian positive definite, this statement further implies that the μ_i are real positive and (31). The latter relation, with statement 1 of the theorem, entails $\max_i (\mu_i^{-1}) \leq \max(\lambda_{\max}(X^{-1}A), 1)$. This, together with $\rho(T) = \max(\max_i (\mu_i^{-1}) - 1, 1 - \min_i (\mu_i^{-1}))$ proves then (32). On the other hand $|1 - \alpha - \lambda| \leq \alpha$ holds if and only if $1 - \lambda$ belongs to the disk of center $(\alpha, 0)$ and radius α ; that is, if and only if $\Re((1 - \lambda)^{-1}) \geq (2\alpha)^{-1}$. With (30), one sees then that (34) holds when

$$\frac{\Re(\mathbf{v}^* A^{-1} \mathbf{v})}{\mathbf{v}^* X^{-1} \mathbf{v}} \geq \frac{1}{2\alpha} \quad \text{for all } \mathbf{v} \in \mathcal{C}^n$$

Further, one has, for all $\mathbf{v} \in \mathcal{C}^n$, and letting $\mathbf{w} = A\mathbf{v}$,

$$\begin{aligned} \frac{\Re(\mathbf{v}^* A^{-1} \mathbf{v})}{\mathbf{v}^* X^{-1} \mathbf{v}} \geq \frac{1}{2\alpha} &\iff \mathbf{v}^* (A^{-1} + (A^*)^{-1}) \mathbf{v} \geq \alpha^{-1} \mathbf{v}^* X^{-1} \mathbf{v} \\ &\iff \mathbf{w}^* (A + A^*) \mathbf{w} \geq \alpha^{-1} \mathbf{w}^* A^* X^{-1} A \mathbf{w} \\ &\iff \mathbf{w}^* ((\alpha I - A^* X^{-1}) X (\alpha I - X^{-1} A) - \alpha^2 X) \mathbf{w} \leq 0 \\ &\iff \|(\alpha I - X^{-1} A) \mathbf{w}\|_X \leq \alpha \|\mathbf{w}\|_X \quad \square \end{aligned}$$

The assumptions of the SPD case referred above imply in fact that X is SPD and that the eigenvalues of $X^{-1}A$ are in the interval $(0, 1]$; (31) and (32) reproduce then the classical results

mentioned above, namely $\sigma(T) \subset [0, 1)$ and $\rho(T) = 1 - (\max_i \mu_i)^{-1}$. Note, however, that our analysis does not require a symmetric (or symmetrizable[§]) smoothing scheme. In particular 1 pre- or post-smoothing step is allowed (e.g. $v_1 = 1$ and $v_2 = 0$). This is seldom used in practice but, as will be seen, allows a nice interpretation of some theoretical results.

Note this interpretation of (31): the eigenvalues of T are bounded by the extremal eigenvalues of $I - X^{-1}A$; that is, when A, X are Hermitian positive definite, the convergence of the two-grid method can never be worse than that of the smoother alone. Roughly speaking, (30) shows that this also holds in the general case if the field of value $\mathbf{v}^* X^{-1} \mathbf{v} / \mathbf{v}^* A^{-1} \mathbf{v}$ excludes the origin and does not exceed much the convex hull of the eigenvalues of $X^{-1}A$. In general, such a relation between the field of value and eigenvalues is difficult to prove. In this respect, the last statement of Corollary 2.1 gives us a sufficient criterion, since the region of the complex plane, which contains the eigenvalues of T according to (34) is also the region where the eigenvalues of $I - X^{-1}A$ are known to be according to (33) (which may be rewritten $\|(1 - \alpha)I - (I - X^{-1}A)\|_X \leq \alpha$). The assumption that X is Hermitian positive definite is restrictive but, as will be seen below, corresponds to an important case from the point of view of the theory of AMG methods. On the other hand, as far as bounding $\rho(T)$ away from 1 is concerned, (34) is of practical interest only if $\alpha < 1$ (and even significantly away from 1). However, this is not a major restriction since

$$\|\alpha I - X^{-1}A\|_X \leq \alpha \iff \|I - (\alpha X)^{-1}A\|_{\alpha X} \leq 1$$

Hence, given a smoother such that $\|I - X^{-1}A\|_X \leq 1$, a proper rescaling suffices to ensure that the condition (33) holds for a nice value of α .

A nice value for α is, for instance, $\alpha = \frac{1}{2}$. Then, the region defined by (34) is the disk of center $(\frac{1}{2}, 0)$ and radius $\frac{1}{2}$. On the other hand, the eigenvalues of T are also necessarily outside the disk of center $(1, 0)$ and radius $(\max_i |\mu_i|)^{-1}$. It then follows from geometric arguments, see Figure 1 (left) that

$$\rho(T) \leq \left(1 - \frac{1}{\max_i |\mu_i|^2}\right)^{1/2} \quad (35)$$

Geometric argument also shows that

$$\rho(T) \leq \left(1 - \min_i \Re\left(\frac{1}{\mu_i}\right)\right)^{1/2} \quad (36)$$

see Figure 1 (right). This allows to exploit a lower bound on $\min_i \Re(\mu_i^{-1})$, which, as will be seen below, sometimes comes as a by product of the analysis of $\max_i |\mu_i|$.

Therefore, Corollary 2.1 allows us to identify cases for which deriving an upper bound on $|\mu_i|$ is sufficient to obtain a rigorous convergence analysis. Unfortunately, there are many cases left aside, for which X is not Hermitian positive definite or (33) does not hold, whereas (30) does not tell too much. Nevertheless, an analysis of $\max_i |\mu_i|$ is relevant in *all cases* since $\rho(T) \geq 1 - (\max_i |\mu_i|)^{-1}$: avoiding large $|\mu_i|$ is always a *necessary* condition for fast convergence. Note also that avoiding large $|\mu_i|$ means that the iteration matrix will have no eigenvalue close to one. If the two-grid

[§]The analyses for the SPD case may be formally applied if, given X , there exists a factorization (7) of $I - X^{-1}A$ satisfying the assumptions of the SPD case; this, in turn, is true if and only if X is SPD and the eigenvalues of $X^{-1}A$ do not exceed 1.

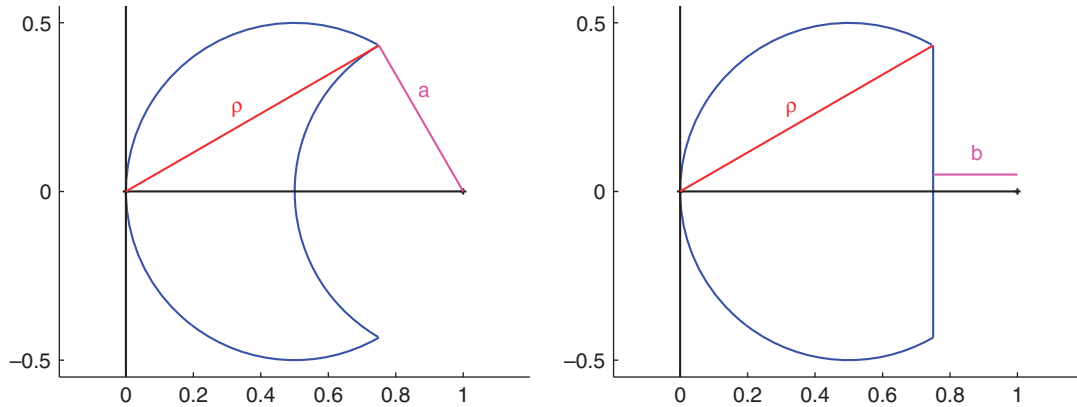


Figure 1. Regions defined by $|\lambda - \frac{1}{2}| \leq \frac{1}{2}$ and: left: $|1 - \lambda| > a$ (since $(0, 0)$, $(1, 0)$ and the intersection of both circles define a right triangle, there holds $\rho^2 + a^2 = 1$); right: $\Re(\lambda) \leq 1 - b$ (since the intersection of the circle and the line $x = 1 - b$ has coordinates $(1 - b, \sqrt{b(1 - b)})$, there holds $\rho^2 = (1 - b)^2 + b(1 - b) = 1 - b$).

method is used as a preconditioner, this is equivalent to the statement: ‘the preconditioned matrix has no eigenvalue close to zero’; and, often, avoiding eigenvalue close to zero for the preconditioned matrix is all what is required for fast convergence of Krylov subspace solvers.

We now discuss how to bound $\max_i |\mu_i|$. Our main results are gathered in the following corollary.

Corollary 2.2

If, in addition to the assumptions of Theorem 2.1, $R = P^*$, then

$$\max_i |\mu_i| \leq \bar{\mu}(A, X) \tag{37}$$

where

$$\bar{\mu}(A, X) = \max_{\mathbf{v} \in \mathbb{C}^n \setminus \{0\}} \frac{|\mathbf{v}^* X (I - \pi_X) \mathbf{v}|}{|\mathbf{v}^* A \mathbf{v}|} \tag{38}$$

Moreover, for any Hermitian positive-definite matrix $Y = L_Y L_Y^*$ and any projector Q onto the range of P

$$\bar{\mu}(A, X) \leq \eta(X, Y) \bar{\mu}_Q(A, Y) \tag{39}$$

where

$$\eta(X, Y) = \|L_Y^{-1} X (I - \pi_X) (L_Y^*)^{-1}\| \tag{40}$$

and where

$$\bar{\mu}_Q(A, Y) = \max_{\mathbf{v} \in \mathbb{C}^n \setminus \{0\}} \frac{\mathbf{v}^* (I - Q)^* Y (I - Q) \mathbf{v}}{|\mathbf{v}^* A \mathbf{v}|} \tag{41}$$

Letting $\pi_Y = P(P^* Y P)^{-1} P^* Y$, one has

$$\bar{\mu}(A, Y) = \bar{\mu}_{\pi_Y}(A, Y) \leq \bar{\mu}_Q(A, Y) \tag{42}$$

i.e. $\bar{\mu}_Q(A, Y)$ is minimal for $Q = \pi_Y$. Letting $A_H = \frac{1}{2}(A + A^*)$ be the Hermitian part of A , there also holds

$$\bar{\mu}_Q(A, Y) = \bar{\mu}_Q(A_H, Y) = \max_{\mathbf{v} \in \mathcal{C}^n \setminus \{0\}} \frac{\mathbf{v}^*(I - Q)^* Y (I - Q) \mathbf{v}}{|\Re(\mathbf{v}^* A \mathbf{v})|} \quad (43)$$

Eventually, if, in addition, X and A_H are Hermitian positive definite,

$$\Re\left(\frac{1}{\mu_i}\right) \geq \frac{1}{\bar{\mu}_Q(A, X)} = \frac{1}{\bar{\mu}_Q(A_H, X)} \quad (44)$$

Proof

Inequality (37) follows from the definition of the μ_i in Theorem 2.1.

Further, $Q = P\tilde{R}$ is a projector onto the range of P if $\tilde{R}P = I$ and, conversely, any such projector can be put in that form. Since $X(I - \pi_X)Q = Q^*X(I - \pi_X) = 0$, (39) follows then from:

$$\begin{aligned} |\mathbf{v}^* X (I - \pi_X) \mathbf{v}| &= |\mathbf{v}^* (I - Q)^* X (I - \pi_X) (I - Q) \mathbf{v}| \\ &\leq \|L_Y^* (I - Q) \mathbf{v}\| \|L_Y^{-1} X (I - \pi_X) (I - Q) \mathbf{v}\| \\ &\leq \|L_Y^{-1} X (I - \pi_X) (L_Y^*)^{-1}\| \|L_Y^* (I - Q) \mathbf{v}\|^2 \end{aligned} \quad (45)$$

Setting $X = Y$ in (45), one also obtains, $\mathbf{v}^* Y (I - \pi_Y) \mathbf{v} \leq \mathbf{v}^* (I - Q)^* Y (I - Q) \mathbf{v}$ (since $\eta(Y, Y) = 1$); this proves (42), whereas (43) follows from:

$$\begin{aligned} \bar{\mu}_Q(A, Y) &\leq \max_{\mathbf{v} \in \mathcal{C}^n \setminus \{0\}} \frac{\mathbf{v}^*(I - Q)^T Y (I - Q) \mathbf{v}}{|\Re(\mathbf{v}^* A \mathbf{v})|} \\ &= \max_{\mathbf{v} \in \mathcal{C}^n \setminus \{0\}} \frac{\mathbf{v}^*(I - Q)^T Y (I - Q) \mathbf{v}}{|\mathbf{v}^* A_H \mathbf{v}|} \quad (= \bar{\mu}_Q(A_H, Y)) \\ &= \max_{\mathbf{v} \in \Re^n \setminus \{0\}} \frac{\mathbf{v}^*(I - Q)^T Y (I - Q) \mathbf{v}}{|\mathbf{v}^* A_H \mathbf{v}|} \\ &= \max_{\mathbf{v} \in \Re^n \setminus \{0\}} \frac{\mathbf{v}^*(I - Q)^T Y (I - Q) \mathbf{v}}{|\mathbf{v}^* A \mathbf{v}|} \\ &\leq \bar{\mu}_Q(A, Y) \end{aligned}$$

Eventually, Statement 3 of Theorem 2.1 (with $P\tilde{R} = Q$ and $\tilde{P}R = Q^*$) yields

$$\mu_i^{-1} = \frac{\mathbf{v}^* A \mathbf{v}}{\mathbf{v}^*(I - Q)^* X (I - \pi_X) (I - Q) \mathbf{v}}$$

for some $\mathbf{v} \in \mathcal{C}^n \setminus \{0\}$. If X is Hermitian positive definite, the denominator on the right-hand side is real, positive and bounded above by $\mathbf{v}^*(I - Q)^* X (I - Q) \mathbf{v}$. Therefore

$$\Re\left(\frac{1}{\mu_i}\right) \geq \min_{\mathbf{v} \in \mathcal{C}^n \setminus \{0\}} \frac{\Re(\mathbf{v}^* A \mathbf{v})}{\mathbf{v}^*(I - Q)^T X (I - Q) \mathbf{v}}$$

and (44) follows because the right-hand side is equal to $\bar{\mu}_Q(A_H, X)^{-1}$ when A_H is positive definite. \square

Note first that our results also apply when A is indefinite. However, $\bar{\mu}(A, X)$ and $\bar{\mu}_Q(A, Y)$ then appear difficult (if not impossible) to bound. This brings another perspective on the well-known difficulty to set up efficient multigrid methods for indefinite problems.

Now we discuss the relationship with previous theories for the SPD case. If A, X are real SPD, the bound (37) reproduces in fact the main result in [10]. There it is further shown that the related bound on the convergence factor is sharp, which can also be seen from our analysis, since (37) is an equality when A, X are Hermitian positive definite, whereas (32) reduces to $\rho(T) = 1 - (\max_i \mu_i)^{-1}$ when $\lambda_{\max}(X^{-1}A)$ is not too large (e.g. does not exceed 1). This sharpness is further exploited in [29] to derive two-side bounds on the convergence factor. The extension of such an approach to nonsymmetric problems seems however difficult because (37) is then no more an equality.

Previous results for the SPD case in [8, 9] are also easily recovered. For $Y = X$ (hence $\eta(X, Y) = 1$), (39), combined with (32), (37), is the bound obtained in [9]; assuming $v_1 = v_2 = 1$ and $M_2 = M_1^T = M^{-1}$, the latter paper further discusses $Y = \frac{1}{2}(M + M^T)$, using implicitly $\eta(X, Y) \leq \lambda_{\max}(Y^{-1}X)$ (see, also [21] for a further analysis of $\lambda_{\max}(Y^{-1}X)$ in this case). On the other hand, with $Y = I$, $v_1 = v_2 = 1$ and $M_2 = M_1 = (\omega/\|A\|)I$ (entailing $\lambda_{\min}(X^{-1}) \geq \omega(2 - \omega)/\|A\|$), one obtains Theorem 3.2 in [8].

Now, whereas (42) shows that the optimal projector is π_Y , the analysis may be easier with simpler choices for Q . For instance, if the two-grid method is based on a Fine/Coarse partitioning and if P has the form

$$P = \begin{pmatrix} J_{FC} \\ I_C \end{pmatrix} \quad (46)$$

where J_{FC} is an interpolation matrix, the most obvious choice for \tilde{R} is

$$\tilde{R} = \begin{pmatrix} 0 & I_C \end{pmatrix} \quad (47)$$

entailing that

$$(I - Q) \begin{pmatrix} \mathbf{v}_F \\ \mathbf{v}_C \end{pmatrix} = \begin{pmatrix} \mathbf{v}_F - J_{FC}\mathbf{v}_C \\ 0 \end{pmatrix}$$

Equation (41) then reduces to

$$\bar{\mu}_Q(A, Y) = \max_{\mathbf{v} \in \mathcal{C}^n \setminus \{0\}} \frac{(\mathbf{v}_F - J_{FC}\mathbf{v}_C)^* Y_{FF} (\mathbf{v}_F - J_{FC}\mathbf{v}_C)}{|\mathbf{v}^* A \mathbf{v}|}$$

where Y_{FF} is the block diagonal part of Y related to the fine-grid variables.

This observation allows to develop the comparison with the ‘standard’ analysis of AMG methods for the SPD case, as developed in [7, 11, 12]. Letting $D = \text{diag}(A)$, the latter is based on a *smoothing property*

$$\|(I - M^{-1}A)\mathbf{v}\|_A^2 \leq \|\mathbf{v}\|_A^2 - \zeta \|\mathbf{v}\|_{AD^{-1}A}^2 \quad \text{for all } \mathbf{v} \in \mathcal{C}^n \quad (48)$$

for some nontrivial (not too small) $\zeta > 0$, and an *approximation property*

$$(\mathbf{v}_F - J_{FC}\mathbf{v}_C)^* D_{FF} (\mathbf{v}_F - J_{FC}\mathbf{v}_C) \leq \tau |\mathbf{v}^* A \mathbf{v}| \quad \text{for all } \mathbf{v} \in \mathcal{C}^n \quad (49)$$

or, more generally,

$$\mathbf{v}^*(I-Q)^*D(I-Q)\mathbf{v} \leq \tau |\mathbf{v}^*A\mathbf{v}| \quad \text{for all } \mathbf{v} \in \mathcal{C}^n \quad (50)$$

for some nontrivial (not too large) τ . Combined, these inequalities imply that $\rho(T) \leq 1 - \zeta/\tau$.

Clearly, $\bar{\mu}_Q(A, D)$ is the best possible constant τ in the approximation property (50), whose extension to the nonsymmetric case has been anticipated by writing $|\mathbf{v}^*A\mathbf{v}|$ instead of $\mathbf{v}^*A\mathbf{v}$ on the right-hand side.

The connection between $\eta(X, D)$ and the smoothing property is apparently less straightforward. However, in the SPD case with $v_1 = v_2 = 1$ and $M_2 = M_1^\top = M^{-1}$, the smoothing property (48) holds if and only if [11, 12]

$$\lambda_{\max}(D^{-1}X) \leq \zeta^{-1}$$

Since $\eta(X, D) \leq \lambda_{\max}(D^{-1}X)$, ζ^{-1} is therefore an upper bound on $\eta(X, D)$. Using this upper bound, (39) combined with (37) and (32) then regives the inequality $\rho(T) \leq 1 - \zeta/\tau$. This connection cannot be extended to the nonsymmetric case, but, at first sight, $\eta(X, D)$ seems not more difficult to analyze directly than the constant ζ in (48).

In fact, smoothers are often well conditioned, hence bounding $\eta(X, D)$ is in general easy compared with the analysis of $\bar{\mu}_Q(A, D)$, which involves ratios whose denominator $|\mathbf{v}^*A\mathbf{v}|$ is small when \mathbf{v} is close to an eigenvector corresponding to an eigenvalue of A that is close to zero.

Thus, the key role is played by the approximation property. In this respect, observe that if $X = \omega^{-1}D$ (for instance, if $M_1 = \omega D^{-1}$ and $v_1 = 1, v_2 = 0$), one has (see (42))

$$\bar{\mu}(A, \omega^{-1}D) = \omega^{-1} \bar{\mu}_{\pi_D}(A, D) \leq \omega^{-1} \bar{\mu}_Q(A, D)$$

Analyzing the smoothing property amounts thus to analyze the two-grid method with only one damped Jacobi pre- or post-smoothing iteration. More precisely, the approximation property (50) provides an upper bound on the convergence factor of this (over) simplified method. Hence, requiring the approximation property to be satisfied for a nontrivial constant τ implies some minimal requirement on the convergence of the two-grid method with only one damped Jacobi smoothing step. This is a sensible requirement for AMG methods which are, in principle, designed to work with simple smoothers.

From a heuristic point of view, this further shows the minor role played by the smoothing property: in some sense, it only ensures that using a more sophisticated smoothing scheme does not deteriorate the convergence.

In the Introduction, we also mention a connection between our results and Hackbush analysis for the SPD case [2, p. 151]. The latter assumes $M_2 = M_1 = M^{-1}$ SPD and is based on the approximation property

$$\mathbf{v}^*(A^{-1} - P A_c^{-1} P^\top)\mathbf{v} \leq c_A \mathbf{v}^* M^{-1} \mathbf{v} \quad \text{for all } \mathbf{v} \in \mathcal{C}^n$$

for some nontrivial constant c_A . The connection is obtained with statement 6 of Theorem 2.1

$$\max_i |\mu_i| \leq \max_{\substack{\mathbf{v} \in \mathcal{C}^n \setminus \{0\} \\ P^* \mathbf{v} = 0}} \frac{|\mathbf{v}^* A^{-1} \mathbf{v}|}{|\mathbf{v}^* X^{-1} \mathbf{v}|} \leq \hat{\mu}(A, X)$$

where

$$\hat{\mu}(A, X) = \max_{\mathbf{v} \in \mathcal{C}^n \setminus \{0\}} \frac{|\mathbf{v}^*(A^{-1} - P A_c^{-1} P^*)\mathbf{v}|}{|\mathbf{v}^* X^{-1} \mathbf{v}|} \tag{51}$$

If A and $M_2 = M_1 = M^{-1}$ are SPD with, for instance, and $v_1 = v_2 = 1$, one has indeed

$$(2 - \lambda_{\max}(M^{-1}A))\mathbf{v}^* M^{-1} \mathbf{v} \leq \mathbf{v}^* X^{-1} \mathbf{v} \leq 2\mathbf{v}^* M^{-1} \mathbf{v}$$

yielding

$$\frac{c_A}{2} \leq \hat{\mu}(A, X) \leq \frac{c_A}{2 - \lambda_{\max}(M^{-1}A)}$$

Note that the bounds in [2, Theorem 6.4.4] are however better than just $\rho(T) \leq 1 - 1/c_A$; further c_A also allows to bound the convergence factor for V- and W-cycle multigrid [2, Theorems 7.2.2 and 7.2.3].

Last but not the least, we conclude the discussion with some comments specific to the nonsymmetric case. First, (43) shows that the analysis of $\bar{\mu}_Q(A, Y)$ reduces to that of $\bar{\mu}_Q(A_H, Y)$; i.e. to the analysis of a symmetric problem. It means that results for the SPD case may be reused in this more general context. This is illustrated in the example below, where we use Theorem A.4.3 in [12], which allows to analyze the approximation property (49) for symmetric M -matrices.

Another interesting observation is that, when X and A_H are Hermitian positive definite, the result (44) is stronger than $|\mu_i| \leq \bar{\mu}_Q(A, X)$ obtained by combining (37) and (39) with $Y = X$ (and hence $\eta(X, X) = 1$). Indeed, $\Re(\mu_i^{-1}) \geq \bar{\mu}_Q(A, X)^{-1}$ implies $|\mu_i| \leq \bar{\mu}_Q(A, X)$ but the converse is not true. Hence, the former inequality restricts more effectively the region allowed to the eigenvalues of the iteration matrix.

2.3. Example

Consider the $n \times n$ matrix

$$A = \begin{pmatrix} 2 & -f(1+p) & & & -f(1-p) \\ -f(1-p) & 2 & -f(1+p) & & \\ & & \ddots & \ddots & \ddots \\ & & & -f(1-p) & 2 & -f(1+p) \\ -f(1+p) & & & -f(1-p) & 2 \end{pmatrix} \tag{52}$$

where $0 < f < 1$ and $|p| \leq 1$. Such matrices arise (up to a scaling factor) when discretizing a convection–diffusion equation $-au''_{xx} + bu'_x + cu = g$ (where $a, c > 0$) on a one-dimensional domain with periodic boundary conditions, using a three-point finite difference scheme on a uniform grid (assuming a form of stabilization if $2a < h|b|$; e.g. enough artificial viscosity).

The eigenvectors of A are the Fourier modes $\mathbf{u}_k, k = 0, \dots, n - 1$ satisfying

$$(\mathbf{u}_k)_l = n^{-1/2} e^{i2kl\pi/n}$$

with corresponding eigenvalue

$$\lambda_k = 2(1 - f \cos(2k\pi/n) + ifp \sin(2k\pi/n))$$

Here, we want to analyze the convergence of the two-grid method with one damped Jacobi smoothing step: $X = 2 \operatorname{diag}(A) = 4I$. We assume n even and that the prolongation corresponds to the standard geometric multigrid; that is it has the form (46), where the coarse variables are the variables with even indexes, and where the interpolation J_{FC} for fine-grid variables corresponds to linear interpolation from neighboring coarse-grid variables (the neighbors of variables with index 1 being the variables with indexes 2 and n). We set $R = P^T$.

We first localize the eigenvalues of T with the help of Corollary 2.1. Any $\mathbf{v} \in \mathcal{C}^n$ may be written in the form $\mathbf{v} = \sum_{k=0}^{n-1} \alpha_k \mathbf{u}_k$, and, since the Fourier modes form an orthonormal basis,

$$\begin{aligned} \frac{\left\| \left(\frac{1}{2}I - X^{-1}A \right) \mathbf{v} \right\|}{\|\mathbf{v}\|} &= \frac{\sum_{k=0}^{n-1} \left| \alpha_k^2 \left(\frac{1}{2} - \frac{\lambda_k}{4} \right) \right|}{\sum_{k=0}^{n-1} |\alpha_k|^2} \\ &\leq \max_k \left| \frac{1}{2} - \frac{\lambda_k}{4} \right| \\ &\leq \frac{f}{2} \end{aligned}$$

Hence (34) holds with $\alpha = \frac{1}{2}$, implying that the eigenvalues of T are in the disk of center $(\frac{1}{2}, 0)$ and radius $\frac{1}{2}$ (this corresponds to the situation depicted in Figure 1).

We now bound $\bar{\mu}(A, X) = \bar{\mu}(A_H, 4I)$, where $A_H = \frac{1}{2}(A + A^T)$; that is, A_H is the matrix (52) for $p=0$. Note that, for any B, C SPD and $0 \leq \alpha \leq \beta$,

$$\bar{\mu}(B + C, \beta I) \leq \max(\bar{\mu}(B, \alpha I), \bar{\mu}(C, (\beta - \alpha)I))$$

Hence, letting

$$B = (1 - f)^{-1}(A_H - 2fI) = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$$

one has

$$\bar{\mu}(A_H, 4I) \leq \max(\bar{\mu}(2fI, 4fI), \bar{\mu}(A_H - 2fI, 4(1 - f)I)) = \max(2, \bar{\mu}(B, 4I))$$

Further, Theorem A.4.3 in [12] proves that the matrix B satisfies the approximation property (49) with a constant τ equal to 1. That is, $\bar{\mu}(B, \operatorname{diag}(B)) = \bar{\mu}(B, 2I) \leq 1$. Thus, there holds

$$\bar{\mu}(A, X) = \bar{\mu}(A_H, 4I) \leq 2$$

and therefore (44) shows that $\sigma(T)$ is also included in the half plane $\Re(z) \leq \frac{1}{2}$ whereas (36) gives

$$\rho(T) \leq \frac{1}{\sqrt{2}}$$

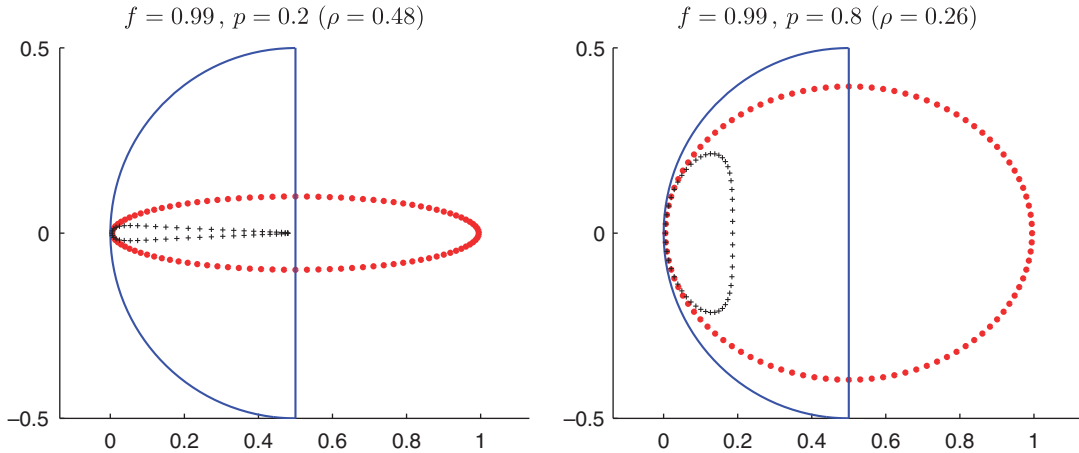


Figure 2. The solid line (—) delimits the region containing $\sigma(T)$ according to our analysis; for $n=100$, the plus signs (+) corresponds to the actual nonzero eigenvalues for T , and the bullets (●) to the eigenvalues of $I - X^{-1}A$.

This is illustrated in Figure 2, where, for $n=100$, we also represent the eigenvalues of $X^{-1}A$ and the nonzero eigenvalues of T , which are here accessible by Fourier analysis:

$$\lambda_k(T) = 1 - \frac{1 + \cos^2(2k\pi/n)}{(1 - \cos(2k\pi/n))^2} + \frac{(1 + \cos(2k\pi/n))^2}{1 - f \cos(2k\pi/n) + ifp \sin(2k\pi/n)} + \frac{(1 + \cos(2k\pi/n))^2}{1 + f \cos(2k\pi/n) - ifp \sin(2k\pi/n)}$$

$$k = 0, \dots, \frac{n}{2} - 1$$

(Note that, despite eigenvalues being known explicitly, it would be hard to prove a further bound on $\rho(T)$.) The results for small p (and further not shown) indicate that our analysis of $\bar{\mu}(A_H, X)$ is accurate: we correctly estimate the largest μ_i for the symmetric problem. However, the bounds (37), (44) are not sharp in this example when the matrix is highly nonsymmetric. From a practical viewpoint, this is a good news: this compensates for the fact that the bounds (35), (36) are less favorable than their counterpart $\rho(T) \leq 1 - (\max_i \mu_i)^{-1}$ for the SPD case (in the considered example, the spectral radius is *smaller* when the matrix is highly nonsymmetric, see Figure 2). From a theoretical point of view, this indicates that a further analysis of $\max_i |\mu_i|$ could be useful. This, however, lies outside the scope of the present paper.

3. COMPATIBLE RELAXATION

The compatible relaxation concept has been introduced in [16] and more thoroughly theoretically justified for the SPD case in [9]. This justification may be summarized as follows. Consider a situation where a Fine/Coarse partitioning has been defined, but $P=R^T$ not necessarily fixed; for instance, there is still some freedom in the interpolation matrix J_{FC} in (46). Then, apply the

smoother alone but restricting the computation to the fine-grid variables, and keep the coarse-grid variables invariant. If the convergence rate of this ‘compatible relaxation’ is fast enough, it is possible to build an interpolation accurate enough (at least theoretically) so that the two-grid method has about the same convergence rate. Otherwise the partitioning has to be revisited. Note that the information retrieved from the compatible relaxation iterations may help to do this, see [17] for practical details.

Hence, compatible relaxation amounts to assessing

$$\rho(I - (S^T M S)^{-1} (S^T A S))$$

where $M^{-1} = M_1$ or $M^{-1} = M_2$, and where

$$S = \begin{pmatrix} I_{n-n_c} \\ 0 \end{pmatrix} \quad (53)$$

is the orthogonal projector onto the fine-grid variables. Here, for the sake of simplicity we consider that one assesses instead

$$\rho_S = \rho(I - (S^T X S)^{-1} (S^T A S))$$

or, equivalently, we consider $\rho(I - (S^T M S)^{-1} (S^T A S))$ but restrict the discussion to schemes with only one smoothing step; this is sensible since convergence estimates derived for this case may be seen, at least from a heuristic point of view, as a worst-case estimates for more general smoothing schemes.

We also restrict the discussion to the real case with $R = P^T$.

Now, letting $A_S = (S^T A S)$ and $X_S = (S^T X S)$,

$$\begin{aligned} \rho_S &= \max_{\lambda \in \sigma(X_S^{-1} A_S)} |1 - \lambda| \\ &\geq 1 - \min_{\lambda \in \sigma(X_S^{-1} A_S)} |\lambda| \end{aligned} \quad (54)$$

$$= 1 - \frac{1}{\rho(A_S^{-1} X_S)} \quad (55)$$

On the other hand, observe that S and $Z = S^T$ fulfill the assumptions in statement 2 of Theorem 2.1, whatever P of the form (46).[‡] Letting

$$C_{S,P} = (S^T A S) - (S^T A P) A_c^{-1} (P^T A S)$$

one has therefore

$$|\mu_i| \leq \rho((C_{S,P}^{-1} A_S)(A_S^{-1} X_S)(X_S^{-1} S^T X (I - \pi_X) S))$$

It is then sensible to expect

$$|\mu_i| \leq c \rho(C_{S,P}^{-1} A_S) \rho(A_S^{-1} X_S) \quad (56)$$

[‡]The approach applies in fact to more general contexts; the discussion to follow requires only that S is an $n \times (n - n_c)$ matrix such that $(S \ P)$ has full rank for the type of prolongation foreseen by the method.

for some nontrivial (not too large) constant c . In fact, when A, X are SPD, (56) holds with $c=1$, whereas the approach is heuristic in other cases.

Now, the key argument in [9] is that, choosing P such that $S^T A P = 0$, one has $C_{S,P} = A_S$, hence (55), (56) provide directly an estimate for $\rho(T)$; for instance, $\rho(T) \leq \rho_S$ if A, X are SPD. Further, a prolongation satisfying this condition is obtained with

$$P = (I - S(S^T A S)^{-1} S^T A) \widehat{R}^T \quad (57)$$

where \widehat{R} is any $n_c \times n$ matrix such that $\widehat{R} S = 0$ and $\widehat{R} \widehat{R}^T = I_{n_c}$. With S given by (53) one may use

$$\widehat{R} = \begin{pmatrix} 0 & I_{n_c} \end{pmatrix}$$

and, letting

$$A = \begin{pmatrix} A_{FF} & A_{FC} \\ A_{CF} & A_{CC} \end{pmatrix}$$

a straightforward computation shows that (57) amounts to

$$P = \begin{pmatrix} -A_{FF}^{-1} A_{FC} \\ I_{n_c} \end{pmatrix}$$

This is compatible with (46) but in most cases impractical since A_{FF}^{-1} is in general a dense matrix.

Fortunately, it is not required to achieve exactly $S^T A P = 0$ to construct relevant P in the SPD case, see, e.g. [30]. From the theoretical point of view, what is needed is $\gamma < 1$, where

$$\gamma^2 = \rho(A_S^{-1} (S^T A P) A_c^{-1} (P^T A S))$$

If $\gamma < 1$, one has indeed $\min_{\lambda \in \sigma(A_S^{-1} C_{S,P})} |\lambda| \geq 1 - \gamma^2$ and (56) gives, using also (55),

$$|\mu_i| \leq \frac{c}{(1 - \gamma^2)(1 - \rho_S)}$$

In the SPD case, γ is known as the Cauchy–Bunyakowski–Schwarz (C.B.S.) constant [31–33] associated with the matrix

$$\widehat{A} = \begin{pmatrix} P^T \\ S^T \end{pmatrix} A \begin{pmatrix} P & S \end{pmatrix} = \begin{pmatrix} P^T A P & P^T A S \\ S^T A P & S^T A S \end{pmatrix} \quad (58)$$

This matrix may be seen as the system matrix A expressed in the generalized hierarchical basis induced by $(P \ S)$ [34]. In this case, it is known that $\gamma < 1$, and γ possesses some nice properties, showing that it measures in fact how well \widehat{A} is approximated by its block diagonal part

$$D_{\widehat{A}} = \begin{pmatrix} P^T A P & \\ & S^T A S \end{pmatrix}$$

The following theorem shows that some of these properties carry over to the nonsymmetric case.

Theorem 3.1

Let

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

be an $n \times n$ matrix with nonsingular square diagonal blocks A_{11} , A_{22} of order n_1 , n_2 , respectively. Assume $n_2 \leq n_1$ and let $\lambda_1, \dots, \lambda_{n_2}$ be the eigenvalues of $A_{22}^{-1}A_{21}A_{11}^{-1}A_{12}$. Let

$$\gamma = \max_{1 \leq i \leq n_2} |\lambda_i|^{1/2}$$

1. $\gamma^2 = \rho(A_{22}^{-1}A_{21}A_{11}^{-1}A_{12}) = \rho(A_{11}^{-1}A_{12}A_{22}^{-1}A_{21})$
2. The eigenvalues of $A_{22}^{-1}(A_{22} - A_{21}A_{11}^{-1}A_{12})$ are $1 - \lambda_1, \dots, 1 - \lambda_{n_2}$.
3. The eigenvalues of $A_{11}^{-1}(A_{11} - A_{12}A_{22}^{-1}A_{21})$ are $1 - \lambda_1, \dots, 1 - \lambda_{n_2}$, plus $n_1 - n_2$ times the eigenvalue 1.
4. Letting

$$D_A = \begin{pmatrix} A_{11} & \\ & A_{22} \end{pmatrix}$$

the eigenvalues of $D_A^{-1}A$ are $1 + \lambda_1^{1/2}, 1 - \lambda_1^{1/2}, \dots, 1 + \lambda_{n_2}^{1/2}, 1 - \lambda_{n_2}^{1/2}$, plus $n_1 - n_2$ times the eigenvalue 1, and

$$\rho(I - D_A^{-1}A) = \gamma$$

Proof

Statements 1–3 are readily proved with the help of Lemma A.1. On the other hand, from

$$\det \left(\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \right) = \det(B_{11}) \det(B_{22} - B_{21}B_{11}^{-1}B_{12})$$

one finds

$$\begin{aligned} \det(D_A^{-1}A - \lambda I) &= \det \left(\begin{pmatrix} (1-\lambda)I_{n_1} & A_{11}^{-1}A_{12} \\ A_{22}^{-1}A_{21} & (1-\lambda)I_{n_2} \end{pmatrix} \right) \\ &= \det((1-\lambda)I_{n_1}) \det((1-\lambda)I_{n_2} - (1-\lambda)^{-1}A_{22}^{-1}A_{21}A_{11}^{-1}A_{12}) \\ &= \det((1-\lambda)I_{n_1}) \det((1-\lambda)^{-1}I_{n_2}) \det((1-\lambda)^2 I_{n_2} - A_{22}^{-1}A_{21}A_{11}^{-1}A_{12}) \\ &= (1-\lambda)^{n_1-n_2} \mathcal{P}((1-\lambda)^2) \end{aligned}$$

where $\mathcal{P}(\lambda)$ is the characteristic polynomial of $A_{22}^{-1}A_{21}A_{11}^{-1}A_{12}$. The fourth statement straightforwardly follows. \square

APPENDIX A

Lemma A.1

Let F be an $n \times m$ matrix and G an $m \times n$ matrix. Let \mathcal{P}_{FG} and \mathcal{P}_{GF} be the characteristic polynomials of FG and GF , respectively. There holds

$$\mathcal{P}_{FG}(\lambda) = \lambda^{n-m} \mathcal{P}_{GF}(\lambda) \quad (\text{A1})$$

Proof

Let

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} xI_n & F \\ G & xI_m \end{pmatrix}$$

where x is any nonzero number. From $\det(B) = \det(B_{11}) \det(B_{22} - B_{21}B_{11}^{-1}B_{12})$ one finds

$$\begin{aligned} \det(B) &= \det(xI_n) \det(xI_m - x^{-1}GF) \\ &= \det(xI_n) \det(x^{-1}I_m) \det(x^2I_m - GF) \\ &= x^{n-m} \mathcal{P}_{GF}(x^2) \end{aligned}$$

Similarly, from $\det(B) = \det(B_{22}) \det(B_{11} - B_{12}B_{22}^{-1}B_{21})$, one obtains

$$\det(B) = \det(xI_m) \det(xI_n - x^{-1}FG) = x^{m-n} \mathcal{P}_{FG}(x^2)$$

Comparing both expressions for $\det(B)$ proves (A1) for all positive λ . Since \mathcal{P}_{FG} and \mathcal{P}_{GF} are polynomials of order n , m , respectively, (A1) then necessarily holds for any λ . \square

ACKNOWLEDGEMENTS

I thank Artem Napov and the referees for their useful comments. The author is supported by the Belgian FNRS ('Directeur de recherches').

REFERENCES

1. Trottenberg U, Oosterlee CW, Schüller A. *Multigrid*. Academic Press: San Diego, London, 2001.
2. Hackbusch W. *Multi-grid Methods and Applications*. Springer: Berlin, 1985.
3. Oswald P. *Multilevel Finite Element Approximation: Theory and Applications*. Teubner Skripte zur Numerik, Teubner: Stuttgart, 1994.
4. Stüben K, Trottenberg KU. Multigrid methods: fundamental algorithms, model problem analysis and applications. In *Multigrid Methods*, Hackbusch W, Trottenberg U (eds). Lectures Notes in Mathematics, vol. 960. Springer: Berlin, Heidelberg, New York, 1982; 1–176.
5. Reusken A. Fourier analysis of a robust multigrid method for convection–diffusion equations. *Numerische Mathematik* 1995; **71**:365–398.
6. Notay Y. A robust algebraic multilevel preconditioner for non symmetric M -matrices. *Numerical Linear Algebra with Applications* 2000; **7**:243–267.
7. Brandt A. Algebraic multigrid theory: the symmetric case. *Applied Mathematics and Computation* 1986; **19**:23–56.
8. Brezina M, Cleary AJ, Falgout RD, Henson VE, Jones JE, Manteuffel TA, McCormick SF, Ruge JW. Algebraic multigrid based on element interpolation (AMGe). *SIAM Journal on Scientific Computing* 2000; **22**:1570–1592.
9. Falgout RD, Vassilevski PS. On generalizing the algebraic multigrid framework. *SIAM Journal on Numerical Analysis* 2005; **42**:1669–1693.

10. Falgout RD, Vassilevski PS, Zikatanov LT. On two-grid convergence estimates. *Numerical Linear Algebra with Applications* 2005; **12**:471–494.
11. Ruge JW, Stüben K. Algebraic multigrid (AMG). In *Multigrid Methods*, McCormick SF (ed.). Frontiers in Applied Mathematics, vol. 3. SIAM: Philadelphia, PA, 1987; 73–130.
12. Stüben K. An introduction to algebraic multigrid. In *Multigrid*, Trottenberg U, Oosterlee CW, Schüller A (eds). Academic Press: San Diego, London, 2001; 413–532. Appendix A.
13. Chartier T, Falgout RD, Henson VE, Jones J, Manteuffel T, McCormick S, Ruge J, Vassilevski PS. Spectral AMG (ρ AMGe). *SIAM Journal on Scientific Computing* 2004; **25**:1–26.
14. Henson VE, Vassilevski PS. Element-free AMG: general algorithms for computing interpolation weights in AMG. *SIAM Journal on Scientific Computing* 2002; **23**:629–650.
15. Vaněk P, Mandel J, Brezina M. Algebraic multigrid based on smoothed aggregation for second and fourth order elliptic problems. *Computing* 1996; **56**:179–196.
16. Brandt A. General highly accurate algebraic coarsening. *Electronic Transactions on Numerical Analysis* 2000; **10**:1–20.
17. Livne OE. Coarsening by compatible relaxation. *Numerical Linear Algebra with Applications* 2004; **11**:205–227.
18. MacLachlan SP, Saad Y. A greedy strategy for coarse-grid selection. *SIAM Journal on Scientific Computing* 2007; **29**:1825–1853.
19. Mense C, Nabben R. On algebraic multilevel methods for non-symmetric systems—comparison results. *Linear Algebra and its Applications* 2008; **429**:2567–2588.
20. Mense C, Nabben R. On algebraic multilevel methods for non-symmetric systems—convergence results. *Electronic Transactions on Numerical Analysis* 2008; **20**:323–345.
21. Notay Y. Algebraic multigrid and algebraic multilevel methods: a theoretical comparison. *Numerical Linear Algebra with Applications* 2005; **12**:419–451.
22. Vassilevski PS. *Multilevel Block Factorization Preconditioners*. Springer: New York, 2008.
23. MacLachlan SP, Oosterlee CW. Algebraic multigrid solvers for complex-valued matrices. *SIAM Journal on Scientific Computing* 2008; **30**:1548–1571.
24. Dendy JE. Black box multigrid for nonsymmetric problems. *Applied Mathematics and Computation* 1983; **13**:261–283.
25. Wienands R, Oosterlee CW. On three-grid Fourier analysis for multigrid. *SIAM Journal on Scientific Computing* 2001; **23**:651–671.
26. Brezina M, Manteuffel T, McCormick S, Ruge J, Sanders G. Towards adaptive smoothed aggregation (α SA) for nonsymmetric problems. *Technical Report*, University of Colorado at Boulder, 2008.
27. Clees T. AMG strategies for PDE systems with applications in industrial semiconductor simulation. *Ph.D. Dissertation*, University of Cologne, Germany, 2004.
28. Notay Y. An aggregation-based algebraic multigrid method. *Technical Report GANMN 08–02*, Université Libre de Bruxelles, Brussels, Belgium, 2008. Available from: <http://homepages.ulb.ac.be/~ynotay>.
29. Zikatanov LT. Two-sided bounds on the convergence rate of two-level methods. *Numerical Linear Algebra with Applications* 2007; **15**:439–454.
30. Brannick J, Zikatanov L. Algebraic multigrid methods based on compatible relaxation and energy minimization. In *Domain Decomposition Methods in Science and Engineering XVI*, Widlund OB, Keyes DE (eds). Lecture Notes in Computational Science and Engineering, vol. 55. Springer: Berlin, Heidelberg, 2007; 15–26.
31. Axelsson O. *Iterative Solution Methods*. Cambridge University Press: Cambridge, 1994.
32. Axelsson O. A survey of algebraic multilevel iterations (AMLI) methods. *BIT* 2003; **43**:863–879.
33. Eijkhout V, Vassilevski PS. The role of the strengthened c.b.s. inequality in multilevel methods. *SIAM Review* 1991; **33**:405–419.
34. Chow E, Vassilevski PS. Multilevel block factorizations in generalized hierarchical bases. *Numerical Linear Algebra with Applications* 2003; **10**:105–127.